AFRL-IF-RS-TR-2006-177
Final Technical Report
May 2006

AN ANALYTICAL FRAMEWORK FOR THE OPTIMAL DESIGN AND DETECTION OF COVERT CHANNEL COMMUNICATIONS

State University of New York @ Albany

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2006-177 has been reviewed and is approved for publication

APPROVED: /s/

STEPHEN J. KUHN
Project Engineer

FOR THE DIRECTOR: /s/

WARREN H. DEBANY, Jr.
Technical Advisor, Information Grid Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	MAY 2006	Final Feb 04 – Feb 06

4. TITLE AND SUBTITLE

AN ANALYTICAL FRAMEWORK FOR THE OPTIMAL DESIGN AND DETECTION OF COVERT CHANNEL COMMUNICATIONS

5. FUNDING NUMBERS

C - FA8750-04-1-0091
PE - 61102F
PR - 2311
TA - 00
WU - 05

6. AUTHOR(S)

Michael J. Medley, Stella N. Batalama and Dimitris A. Pados

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

State University of New York @ Albany
35 State Street
Albany NY 12207-2916

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/IFGB
525 Brooks Road
Rome NY 13441-4505

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-2006-177

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Stephen J. Kuhn, IFGB, Stephen.Kuhn@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; distribution unlimited. PA# 06-353

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

We propose an iterative generalized least squares procedure to recover unknown messages hidden in image hosts via spread-spectrum embedding. Neither the original host nor the embedding signature is assumed available. We demonstrate that for hidden messages of sufficient length (data sample support), recovery can be achieved with probability of error close to what may be attained with known embedding signature and known original host autocorrelation matrix. For small hidden messages, the signature estimate calculated by the iterative generalized least squares procedure can be fed as initial value to a (computationally costly) expectation-maximization signature identification scheme that we derive. Message recovery can again be carried out successfully by means of a linear sample-matrix-inversion minimum-mean-square-error receiver.

14. SUBJECT TERMS

Capacity, distortion, linear filters, signal-to-interference-plus-noise ratio, spread spectrum, Steganography, watermarking

15. NUMBER OF PAGES

24

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Contents

List of Figures

I. Introduction

Spread-spectrum (SS) steganography is attracting increasing interest among researchers and practitioners in the fields of authentication and covert communications. Under a blind host medium scenario, that is when the original clean host is assumed to be unavailable, past works focused on the detection of the presence of a known message as in [1], [2], the recovery of an unknown message embedded with a known signature as in [3], [4], or system optimization for covert message delivery to signature-aware recipients [5]-[7].

Yet, one challenging issue in SS data hiding applications is fully blind message recovery, that is when little or nothing can be assumed about the embedded message, the embedding signature, and the host image, with direct application to eavesdropping the communication of an enemy. Moreover, fully blind message recovery is also of interest when the communication takes place between allies, since it is bandwidth efficient and requires no prior agreement on the message to be sent, its signature, the stego-key, or a training sequence.

In a single or superimposed signal in additive white Gaussian noise (AWGN) scenario one might split the two subspaces spanned by the signal and the noise eigenvectors, respectively [18]. However, when the message of interest and its total disturbance lie in the same subspace and the message signal components cannot be distinguished by subspace estimation techniques, or when subspace tracking is dubious, blind signal separation and estimation is usually achieved through iterative procedures that take advantage of the finite alphabet property arising from the digital nature of the hidden data and—when applied to multiple input channels—the independence of the signals to be separated.

Throughout the course of our research, we attempt to recover a hidden unknown message when neither the original host nor the embedding signature is known (fully blind SS steganalysis). The only prior knowledge assumed (or guessed) is the embedding (block) transform domain. In blind SS steganalysis the unknown host image acts as a source of interference to the message to be extracted and, in a way, the problem parallels blind digital signal separation applications as they arise in the fields of array processing [8]–[12], biomedical signal processing [13], [14], image reconstruction [15], [16] or code-division-multiple-access (CDMA) wireless communication systems [17]-[21].

From this point of view, we first develop a least-squares-type iterative procedure for coupled signature estimation and message recovery. Next, under a (colored) Gaussian assumption on the host data bins we are able to treat the message signature as an unknown vector parameter of a Gaussian mixture and derive an iterative procedure for signature-only estimation based on expectation-maximization (EM) principles. Message recovery is then accomplished via minimum-mean-square-error (MMSE) filtering and detection.

The rest of the paper is organized as follows. In Section II we present the received signal model. Iterative procedures based on least squares and EM concepts for signature identification and message recovery are presented in Section III. Simulation results are presented in Section IV. Finally, some conclusions are drawn in Section V.

II. Signal Model and Notation

Basic Notation and Terms

To understand and research the problem of SS steganalysis, it helps to first put ourselves on the side of the SS steganographer. Consider a host image $\mathbf{H} \in \mathcal{A}^{N_1 \times N_2}$ that is to be watermarked where $\mathcal{A}$ is the image alphabet and $N_1 \times N_2$ is the image size in pixels. Fig. 1 shows a gray scale F-16 image example in $\mathcal{A}^{N_1 \times N_2} = \{0, 1, \cdots, 255\}^{512 \times 512}$. Without loss of generality, the image $\mathbf{H}$ is divided into M local blocks of size $\frac{N_1 \times N_2}{M}$ pixels. Each block $\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_M$ is to carry one hidden bit $d(m) \in \{\pm 1\}$, $m = 1, 2, \cdots, M$, respectively. Embedding is performed in a real 2-dimensional transform domain $\mathcal{T}$. After transform calculation and conventional zig-zag scanning vectorization, we obtain $\mathcal{T}(\mathbf{H}_m) \in \mathbb{R}^{\frac{N_1 \times N_2}{M}}$, $m = 1, 2, \cdots, M$. From the transform domain vectors $\mathcal{T}(\mathbf{H}_m)$ we choose a fixed subset of $L \leq \frac{N_1 \times N_2}{M}$ coefficients (bins) to form the final host vectors $\mathbf{x}_m \in \mathbb{R}^L$, $m = 1, 2, \ldots, M$ (for example, it is common and appropriate to exclude the dc coefficient).

An important statistical quantity for the developments that follow is the autocorrelation matrix of the host data $\mathbf{x}$,

$$\mathbf{R}_x \stackrel{\triangle}{=} E\left\{\mathbf{x}\mathbf{x}^T\right\} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^T \tag{1}$$

where $E\{\cdot\}$ denotes statistical expectation (here with respect to $\mathbf{x}$ over the given image $\mathbf{H}$) and T is the transpose operator. It is easy to verify [6],[7] that, in general, $\mathbf{R}_x \neq c\mathbf{I}_L$, $c > 0$, where $\mathbf{I}_L$ is the size-L identity matrix. That is, $\mathbf{R}_x$ *is not* constant-value diagonal or "white" in field language. As a corroborative, illustrative example of this assertion, Fig. 2(a) shows the familiar baboon gray-scale image from standard image processing/benchmarking databases with $\mathbf{H} \in \{0, 1, \cdots, 255\}^{256\times 256}$; Fig. 2(b) shows the baboon host autocorrelation matrix for block-DCT SS embedding of $M = 1,024$ bits (one bit per 8×8 block) with $L = 63$ (embedding over all frequency bins except dc).

In SS steganography, the message bit sequence $\{d(m)\}_{m=1}^{M}$ is hidden in the transform-domain host vectors $\{\mathbf{x}_m\}_{m=1}^{M}$ via additive SS embedding by means of a signature $\mathbf{v} \in \mathbb{R}^L$:

$$\mathbf{y}_m = d(m)\mathbf{v} + \mathbf{x}_m \;\; m = 1, \ldots, M. \tag{2}$$

At this point, it is advantageous to change our perspective to that of the steganalysis researcher attempting to recover the covert message. It is reasonable to assume that $\{d(m)\}_{m=1}^{M}$ behave as equiprobable binary random variables that are independent from each other and $\{\mathbf{x}_m\}_{m=1}^{M}$. It is also deemed reasonable to treat $\{\mathbf{x}_m\}_{m=1}^{M}$ as a sequence of independent identically distributed random vectors with zero mean. Hence, the watermarked data vectors $\{\mathbf{y}_m\}_{m=1}^{M}$ have mean zero, autocorrelation matrix $\mathbf{R}_y = E\left\{\mathbf{y}_m\mathbf{y}_m^T\right\} = \mathbf{R}_x + \mathbf{v}\mathbf{v}^T$, and joint probability distribution function (pdf) $f_{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_M}(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M) = f_{\mathbf{y}_1}(\mathbf{y}_1)f_{\mathbf{y}_2}(\mathbf{y}_2)\ldots f_{\mathbf{y}_M}(\mathbf{y}_M)$. The mean squared distortion of the original image due to the hidden message is

$$\mathcal{D} = E\left\{\|d(m)\mathbf{v} + \mathbf{x}_m - \mathbf{x}_m\|^2\right\} = \|\mathbf{v}\|^2. \tag{3}$$

Basic Detection of Hidden Bits

We are interested in detecting the hidden bits $\{d(m)\}_{m=1}^{M}$ in (2). The linear filter that operates on $\mathbf{y}_m$ and minimizes the mean square error (MSE) at its output is

$$\mathbf{w}_{\text{MMSE}} = \arg\min_{\mathbf{w}} E\left\{\left\|\mathbf{w}^T\mathbf{y}_m - d(m)\right\|^2\right\} = \mathbf{R}_y^{-1}\mathbf{v}. \tag{4}$$

If $\mathbf{v}$ were known, then bit detection could be carried out by sign detection at the MMSE filter output:

$$\hat{d}(m) = \text{sgn}\left(\mathbf{w}_{\text{MMSE}}^T\mathbf{y}_m\right) = \text{sgn}\left(\mathbf{v}^T\mathbf{R}_y^{-1}\mathbf{y}_m\right). \tag{5}$$

Yet, $\mathbf{v}$ is assumed to be unavailable herein. In the sequel, we develop two iterative procedures for signature estimation and message bit recovery. The first procedure is least squares driven and couples a signature estimation and a bit detection step. The second procedure is derived from the EM theory and attempts signature-only identification; message recovery can then be carried out separately.

III. Iterative Procedures

Iterative Generalized Least Squares

For notational simplicity we form the compound data observation matrix

$$\mathbf{Y} = \mathbf{v}\mathbf{d}^T + \mathbf{X} \tag{6}$$

where $\mathbf{d} \stackrel{\triangle}{=} [d(1)\ d(2)\ \dots\ d(M)]^T \in \{\pm 1\}^M$ and $\mathbf{X} \stackrel{\triangle}{=} [\mathbf{x}_1\, \mathbf{x}_2\, \dots\, \mathbf{x}_M] \in \mathbb{R}^{L\times M}$. After data prewhitening, the generalized least squares estimator of $\mathbf{d}$, $\mathbf{v}$ is given by

$$\arg \min_{\mathbf{v}\in\mathbb{R}^L, \mathbf{d}\in\{\pm 1\}^M} \left\| \mathbf{R}_x^{-\frac{1}{2}} \left(\mathbf{Y} - \mathbf{v}\mathbf{d}^T\right) \right\|_F^2 \tag{7}$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm. Notice that if we were allowed to assume that $\{\mathbf{x}_m\}_{m=1}^M$ were Gaussian, then (7) would coincide with maximum likelihood (ML) joint estimation of $\mathbf{d}$, $\mathbf{v}$ (treating both $\mathbf{d}$ and $\mathbf{v}$ as deterministic unknown parameters). In any case, regretfully, joint estimation of $\mathbf{d}$, $\mathbf{v}$ by (7) has complexity exponential in the hidden message length M and event the shortest of hidden messages, say 100 bits, makes the recovery task practically impossible. As such, we consider this cost unacceptable and attempt to reach a quality approximation of the solution by alternating least squares estimates of $\mathbf{d}$, $\mathbf{v}$ iteratively, as follows:

1. Pretend $\mathbf{v}$ is known;

   · Then, the least squares (LS) estimate of $\mathbf{d}$ is

$$\hat{\mathbf{d}}_{\mathrm{LS}} = \arg \min_{\mathbf{d}\in\{\pm 1\}^M} \left\| \mathbf{R}_x^{-\frac{1}{2}} \left(\mathbf{Y} - \mathbf{v}\mathbf{d}^T\right) \right\|_F^2 = \mathrm{sgn}(\mathbf{Y}^T\mathbf{R}_x^{-1}\mathbf{v}). \tag{8}$$

· Observing that $\mathbf{R}_x^{-1}\mathbf{v} = c\mathbf{R}_y^{-1}\mathbf{v}$, $c > 0$, we rewrite

$$\hat{\mathbf{d}}_{\text{LS}} = \text{sgn}(\mathbf{Y}^T\mathbf{R}_y^{-1}\mathbf{v}) \tag{9}$$

and recognize that (9) represents MMSE filtering followed by sign detection.

2. Pretend, in turn, that $\mathbf{d}$ is known;

· Then the least squares estimate of $\mathbf{v}$ is

$$\hat{\mathbf{v}}_{\text{LS}} = \arg\min_{\mathbf{v}\in\mathbb{R}^L} \left\| \mathbf{R}_x^{-\frac{1}{2}} \left(\mathbf{Y} - \mathbf{v}\mathbf{d}^T\right) \right\|_F^2 = \frac{1}{M}\mathbf{Y}\mathbf{d}. \tag{10}$$

The conditional LS estimate in (10) is the best—minimum variance—unbiased estimate of $\mathbf{v}$ [22].

The proposed algorithm is now straightforward. Initialize $\hat{\mathbf{v}}$ (or $\hat{\mathbf{d}}$) arbitrarily (or by an educated guess if side information is available) and alternate iteratively between (9) and (10) to obtain at each step conditionally least squares optimal estimates of one vector parameter given the other. Stop when convergence is observed. Notice that (9) requires knowledge of the autocorrelation matrix of the watermarked host data $\mathbf{R}_y$ which can be estimated by sample averaging over the received data observations, $\hat{\mathbf{R}}_y(M) = \frac{1}{M}\sum_{m=1}^{M}\mathbf{y}_m\mathbf{y}_m^T$. Of course, this way the optimality—in the LS sense—of (9) is obtained asymptotically, since $\hat{\mathbf{R}}_y(M) \underset{M\to\infty}{\to} \mathbf{R}_y$ in probability for elliptically contoured input vectors [23]. We call the coupled repeated calculation of (9) and (10) iterative generalized least squares[1] SS steganalysis (IGLS-SSS).

Extensive experimentation with (9) and (10) showed that for sufficiently long hidden messages (number of bits M equal to 4,000 or more, for example) high quality message decisions $\hat{\mathbf{d}}$ are obtained. When the message size is small, however, estimation of $\mathbf{R}_y$ becomes problematic causing deviations of the least squares estimates of $\hat{\mathbf{d}}$ (and $\hat{\mathbf{v}}$) from the true values. To address this concern, we have considered the possibility of developing an expectation- maximization (EM) signature identification procedure as a final-stage assist.

Although in general EM schemes can be computationally expensive and slow in convergence, meaningful improvements for small hidden messages have been achieved, as seen in [24], and can be exploited. In fact, in the following section, we derive a new expectation-maximization (EM) signature identification procedure to handle the aforementioned cases.

[1]Occasionally, conditional coupled least-squares schemes as the one in (9) and (10) are referred to as weighted least squares [13] in the literature.

Expectation-Maximization

We are interested in estimating $\mathbf{v}$ from the received data vector $\mathbf{y} \stackrel{\triangle}{=} [\mathbf{y}_1^T \ \mathbf{y}_2^T \ \cdots \ \mathbf{y}_M^T]^T \in \mathbb{R}^{LM}$ with pdf $f_{\mathbf{y}}(\mathbf{y};\mathbf{v})$. To overcome the computational intractability of the maximum likelihood estimate of $\mathbf{v}$, $\hat{\mathbf{v}}_{\mathrm{ML}} = \arg\max_{\mathbf{v}\in\mathbb{R}^L} f_{\mathbf{y}}(\mathbf{y};\mathbf{v})$, we view the received data $\mathbf{y}$ as a set of incomplete data that are part of a larger postulated dataset of complete data $[\mathbf{y},\mathbf{d}]$. The pdf of the incomplete data (parameterized in $\mathbf{v}$) is given by

$$f_{\mathbf{y}}(\mathbf{y};\mathbf{v}) = \sum_{\mathbf{d}\in\{\pm 1\}^M} f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v}) \tag{11}$$

where $f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v})$ denotes the joint pdf of $\mathbf{y},\mathbf{d}$ parameterized in $\mathbf{v}$. If $\mathbf{d}$ were available, then the pair $(\mathbf{y},\mathbf{d})$ would be a sufficient statistic for ML estimation of $\mathbf{v}$ by $\hat{\mathbf{v}}_{\mathrm{ML}} = \arg\max_{\mathbf{v}\in\mathbb{R}^L} f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v}) = \arg\max_{\mathbf{v}\in\mathbb{R}^L}\{\ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v})\}$. Since $\mathbf{d}$ is not available, we may replace/estimate $\ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v})$ by a function, say $Q(\mathbf{v})$, $Q(\mathbf{v}) \approx \ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v})$ and maximize that estimated function, i.e.

$$\hat{\mathbf{v}} = \arg\max_{\mathbf{v}\in\mathbb{R}^L} Q(\mathbf{v}). \tag{12}$$

To proceed further we select as $Q(\mathbf{v})$ the MMSE estimate of $\ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v})$ given $\mathbf{y}$. Therefore, $Q(\mathbf{v}) = E_{\mathbf{d}}\{\ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v})|\mathbf{y}\}$. However, the conditional expectation $E_{\mathbf{d}}\{\ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v})|\mathbf{y}\}$ includes the conditional pmf of the hidden data $p_{\mathbf{d}/\mathbf{y}}(\mathbf{d})$ for all $\mathbf{d}\in\{\pm 1\}^M$, which in turn depends on $\mathbf{v}$, hence is not available. If an estimate $\hat{\mathbf{v}}^{(k)}$ is available, it can be used to define the conditional pmf of the hidden data parameterized in $\hat{\mathbf{v}}^{(k)}$, $p_{\mathbf{d}/\mathbf{y}}(\mathbf{d};\hat{\mathbf{v}}^{(k)})$. Consequently, our estimated function (approximation of $\ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v})$) becomes

$$\begin{aligned} Q(\mathbf{v};\hat{\mathbf{v}}^{(k)}) &\stackrel{\triangle}{=} E_{\mathbf{d}}\left\{\ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v})|\mathbf{y};\hat{\mathbf{v}}^{(k)}\right\} \\ &= \sum_{\mathbf{d}\in\{\pm 1\}^M} p_{\mathbf{d}/\mathbf{y}}(\mathbf{d};\hat{\mathbf{v}}^{(k)}) \ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y},\mathbf{d};\mathbf{v}) \end{aligned} \tag{13}$$

and (12) becomes $\hat{\mathbf{v}} = \arg\max_{\mathbf{v}\in\mathbb{R}^L} Q(\mathbf{v};\hat{\mathbf{v}}^{(k)})$. This $\hat{\mathbf{v}}$ can be used to obtain an updated estimate

$$\hat{\mathbf{v}}^{(k+1)} = \arg\max_{\mathbf{v}\in\mathbb{R}^L} Q(\mathbf{v};\hat{\mathbf{v}}^{(k)}). \tag{14}$$

Alternating calculation of (13) and (14) constitutes an expectation-maximization (EM) procedure with "objective function" (as frequently referred to in the literature [25]) $Q(\mathbf{v};\hat{\mathbf{v}}^{(k)})$. To

proceed further toward a specific solution to our steganalysis (unknown signature estimation) problem, we choose to model the host data vectors $\mathbf{x}_m$ as Gaussian distributed, $\mathbf{x}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$, $m = 1, 2, \ldots, M$, and therefore

$$f_{\mathbf{y}}(\mathbf{y}; \mathbf{v}) = \left(\frac{1}{2\sqrt{(2\pi)^L |\mathbf{R}_x|}} \right)^M \prod_{m=1}^{M} \left\{ e^{-(\mathbf{y}_m - \mathbf{v})^T \mathbf{R}_x^{-1} (\mathbf{y}_m - \mathbf{v})/2} + e^{-(\mathbf{y}_m + \mathbf{v})^T \mathbf{R}_x^{-1} (\mathbf{y}_m + \mathbf{v})/2} \right\} \tag{15}$$

We rewrite (13) as

$$Q(\mathbf{v}; \hat{\mathbf{v}}^{(k)}) = \sum_{i=1}^{2^M} p_{\mathbf{d}/\mathbf{y}}(\mathbf{d}_i; \hat{\mathbf{v}}^{(k)}) \ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y}, \mathbf{d}_i; \mathbf{v}) \tag{16}$$

where $\mathbf{d}_i \in \{\pm 1\}^M$ is the ith possible message bit combination, $i = 1, 2, \ldots, 2^M$.

Eliminating the constant over i terms in the lefthandside pdf of (16) we obtain

$$\begin{aligned} \ln f_{\mathbf{y},\mathbf{d}}(\mathbf{y}, \mathbf{d}_i; \mathbf{v}) &= \ln f_{\mathbf{y}/\mathbf{d}}(\mathbf{y}; \mathbf{v}) + \ln p_{\mathbf{d}}(\mathbf{d}_i) \\ &= \sum_{m=1}^{M} \mathbf{v}^T \mathbf{R}_x^{-1} \mathbf{y}_m d_i(m) - \frac{M}{2} \mathbf{v}^T \mathbf{R}_x^{-1} \mathbf{v} + c_1 \end{aligned} \tag{17}$$

where $d_i(m)$ denotes the $m-$th element of vector $\mathbf{d}_i \in \{\pm 1\}^M$ and c_1 is a constant that can be dropped as inconsequential to the optimization problem. Substituting (17) in (16) yields

$$\begin{aligned} &Q(\mathbf{v}; \hat{\mathbf{v}}^{(k)}) \\ &= \sum_{i=1}^{2^M} p_{\mathbf{d}/\mathbf{y}}(\mathbf{d}_i; \hat{\mathbf{v}}^{(k)}) \left\{ \sum_{m=1}^{M} \mathbf{v}^T \mathbf{R}_x^{-1} \mathbf{y}_m d_i(m) - \frac{M}{2} \mathbf{v}^T \mathbf{R}_x^{-1} \mathbf{v} + c_1 \right\} \\ &= \sum_{m=1}^{M} \sum_{d=\pm 1} p_{d(m)/\mathbf{y}_m}(d; \hat{\mathbf{v}}^{(k)}) \mathbf{v}^T \mathbf{R}_x^{-1} \mathbf{y}_m d - \frac{M}{2} \mathbf{v}^T \mathbf{R}_x^{-1} \mathbf{v} + c_1. \end{aligned} \tag{18}$$

Let us now calculate the conditional probability

$$\begin{aligned} p_{d(m)/\mathbf{y}_m}(d; \hat{\mathbf{v}}^{(k)}) &= \frac{f_{d(m),\mathbf{y}_m}(d, \mathbf{y}_m; \hat{\mathbf{v}}^{(k)})}{f_{\mathbf{y}_m}(\mathbf{y}_m; \hat{\mathbf{v}}^{(k)})} \\ &= \frac{f_{\mathbf{y}_m/d(m)}(\mathbf{y}_m; \hat{\mathbf{v}}^{(k)})}{\sum_{d=\pm 1} f_{\mathbf{y}_m/d(m)}(\mathbf{y}_m; \hat{\mathbf{v}}^{(k)})}. \end{aligned} \tag{19}$$

Combining (18) and (19), (16) takes the closed form

$$\begin{aligned} Q(\mathbf{v};\hat{\mathbf{v}}^{(k)}) =& \sum_{m=1}^{M} \frac{\sum_{d=\pm 1} f_{\mathbf{y}_m/d(m)}(\mathbf{y}_m;\hat{\mathbf{v}}^{(k)})\mathbf{v}^T\mathbf{R}_x^{-1}\mathbf{y}_m d}{\sum_{d=\pm 1} f_{\mathbf{y}_m/d(m)}(\mathbf{y}_m;\hat{\mathbf{v}}^{(k)})} \\ & - \frac{M}{2}\mathbf{v}^T\mathbf{R}_x^{-1}\mathbf{v} + c_1 \\ =& \sum_{m=1}^{M} \mathbf{v}^T\mathbf{R}_x^{-1}\mathbf{y}_m \frac{(1-e^{-2\hat{\mathbf{v}}^{(k)T}\mathbf{R}_x^{-1}\mathbf{y}_m})}{(1+e^{-2\hat{\mathbf{v}}^{(k)T}\mathbf{R}_x^{-1}\mathbf{y}_m})} \\ & - \frac{M}{2}\mathbf{v}^T\mathbf{R}_x^{-1}\mathbf{v} + c_1 \\ =& \mathbf{v}^T\mathbf{R}_x^{-1}\sum_{m=1}^{M}\mathbf{y}_m \tanh(\hat{\mathbf{v}}^{(k)T}\mathbf{R}_x^{-1}\mathbf{y}_m) \\ & - \frac{M}{2}\mathbf{v}^T\mathbf{R}_x^{-1}\mathbf{v} + c_1. \end{aligned} \tag{20}$$

Conveniently, (20) is quadratic in $\mathbf{v}$. To update the estimate of $\mathbf{v}$, we set the derivative of (20) with respect to $\mathbf{v}$ to zero and obtain the recursive formula

$$\hat{\mathbf{v}}^{(k+1)} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{y}_m \tanh(\hat{\mathbf{v}}^{(k)T}\mathbf{R}_x^{-1}\mathbf{y}_m). \tag{21}$$

In summary, recursion (21) is the proposed EM-type signature identification procedure. Success of the EM recursion in finding the global maximum depends, in general, on the initialization vector $\hat{\mathbf{v}}^{(0)}$. As with the IGLS procedure, $\mathbf{R}_x$ in (21) is substituted by $\hat{\mathbf{R}}_y(M)$.

Discussion on proposed blind IGLS and EM procedures

It is well understood that in comparison with least squares algorithms, EM procedures converge slowly. For the specifics of our steganalysis problem where the autocorrelation matrix of the host data $\mathbf{R}_x$ is not available and is substituted by the sample average estimate of the received data $\hat{\mathbf{R}}_y(M)$, we observed experimentally that the expected likelihood increments in (21) become exceedingly small resulting in uncomfortably slow convergence of the EM procedure. As a result, we do not consider the derived EM recursion as a stand-alone steganalysis scheme. We rather suggest use of (21) as a potential add-on to IGLS for small-sample-support steganalysis initializing $\hat{\mathbf{v}}^{(0)}$ at $\hat{\mathbf{v}}_{\mathrm{LS}}$ from (9), (10).

Finally, for the sake of mathematical accuracy we should emphasize that there is always a phase/sign ambiguity present when one considers joint data demodulation and signature identification. The ambiguity problem can be overcome either under differential data embedding or with a few known (or guessed) embedded data symbols for phase/sign correction. In the following experimental studies the ambiguity problem is assumed handled.

IV. Experimental Studies

We consider as a host example the familiar gray scale 512×512 "F-16 Aircraft" image that has been used widely in the pertinent literature. We perform 8×8 block DCT embedding by (2) over all bins except the dc coefficient with an arbitrary signature $\mathbf{v} \in \mathbb{R}^{63}$ and varying host distortion $\mathcal{D} = \|\mathbf{v}\|^2$. The hidden message is, therefore, $\frac{512^2}{8^2} = 4,096$ bits long. We examine four different message recovery operations: *(i)* Standard signature matched-filtering (MF) with known $\mathbf{v}$; *(ii)* MMSE filtering with known $\mathbf{v}$ and known true autocorrelation matrix of the host $\mathbf{R}_x$ which serves as a performance bound reference for the proposed blind schemes; *(iii)* blind (neither $\mathbf{v}$ nor $\mathbf{R}_x$ is known) IGLS by (9), (10); and *(iv)* blind IGLS followed by EM in (21) as derived and discussed in Section 3 for potentially improved signature identification under small-sample-support steganalysis. Fig. 3 shows the corresponding probability of error (bit-error-rate or BER) curves as a function of the host distortion. To our satisfaction, the proposed all-blind schemes vastly outperform MF recovery with a known embedding signature and approach rather closely the performance of the ideal MMSE detector where both the embedding signature and the host autocorrelation matrix are perfectly known[2]. It is seen that for this given message size (4,096 bits) the EM add-on to IGLS offers a small gain that arguably does not justify the significant increase in computational complexity and decision delay.

In Fig. 4, however, we repeat the exact same study for the small 280×280 version of the Aircraft image. It can now be argued that for this sample support ($\frac{280^2}{8^2} = 1,225$ message bits) the gain of EM post-processing (close to an order of magnitude at 30 dB distortion) does justify

[2]The ideal MMSE detector can be viewed as the receiver of the intended recipient of the hidden message that has knowledge of the signature and the clean host.

the extra cost. (Fig. 5 shows the clean and stego image after message embedding with 30 dB distortion). For this sample support ($\frac{280^2}{8^2} = 1,225$ message bits), the gain of EM post-processing (close to an order of magnitude at 30 dB distortion) is appealing but it is expected that this can be improved if auxiliary-vector (AV) filtering is incorporated either alone or in conjunction with EM. For additional experimental verification, the same study is carried out (Fig. 6) on the 280×280 "Lena" image with the same conclusions.

Finally, Fig. 7 shows an example with the ASCII text message hidden in Fig. 5(b) (with 26 dB image distortion) and its recovered form by IGLS alone or IGLS and EM in tandem.

V. Conclusions

We considered the problem of recovering a hidden message embedded in an unknown digital host image by means of an unknown signature. We first developed an iterative generalized least squares (IGLS) procedure that allows joint signature estimation and message recovery. For large data support (i.e. large images) the hidden message can be blindly recovered with probability of error close to that obtained via supervised MMSE detection. To handle the cases of small data support (i.e. small images) we derived an expectation-maximization procedure that—when initialized appropriately—yields improved probability of error rates when compared to our first proposed scheme. Despite any difficulties associated with the estimation of the image autocorrelation matrix and the subsequent existence of fixed points that are not global solutions, the procedures developed herein provide a computationally feasible alternative to complete enumeration (for signature estimation and/or bit detection).

References

[1] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of nonadditive watermarks," *IEEE Trans. Image Proc.*, vol. 10, pp. 755-766, May 2001.

[2] C. Qiang and T. S. Huang, "Robust optimum detection of transform domain multiplicative watermarks," *IEEE Trans. Signal Proc.*, vol. 51, pp. 906-924, Apr. 2003.

[3] J. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Proc.*, vol. 9, pp. 55-68, Jan. 2000.

[4] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Optimum decoding and detection of multiplicative watermarks," *IEEE Trans. Signal Proc.*, vol. 51, pp. 1118-1123, Apr. 2003.

[5] H. S. Malvar and D. A. Florêncio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Proc.*, vol. 51, pp. 898-905, Apr. 2003.

[6] M. Gkizeli, D. A. Pados, and M. J. Medley, "SINR, bit error rate, and Shannon capacity optimized spread-spectrum steganography," in *Proc. IEEE Intern. Conf. Image Proc.*, Singapore, Oct. 2004, vol. 2, pp. 1561-1564.

[7] M. Gkizeli, D. A. Pados, and M. J. Medley, "Optimal signature design for spread-spectrum steganography," submitted to *IEEE Trans. Image Proc.*.

[8] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust. Speech Signal Proc.*, vol. 36, pp. 477-489, Apr. 1988.

[9] M. Viberg and M. Ottersten, "Sensor array processing based on subspace fitting ," *IEEE Trans. Signal Proc.*, vol. 39, pp. 1110-1121, May. 1991.

[10] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array-part I: algorithms," *IEEE Trans. Signal Proc.*, vol. 44, pp. :1184-1197, May 1996.

[11] A. Ranheim, "A decoupled approach to adaptive signal separation using an antenna array," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 676-682, May 1999.

[12] L. Tao and N. D. Sidiropoulos, "Blind digital signal separation using successive interference cancellation iterative least squares," *IEEE Trans. Signal Proc.*, vol. 48, pp. 3146-3152, Nov. 2000.

[13] J. M. M. Anderson, B. A. Mair, M. Rao, and C. -H. Wu, "Weighted least-squares reconstruction methods for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 16, pp. 159-165, Apr. 1997.

[14] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum a "posteriori" estimates in emission tomography," *IEEE Trans. Med. Imag.* vol. 20, pp. 280-288, Apr. 2001.

[15] A. K. Katsaggelos and K. T. Lay, "Maximum likelihood blur identification and image restoration using the EM algorithm," *IEEE Trans. Signal Proc.*, vol. 39, pp. 729-733, Mar. 1991.

[16] M. A. T. Figueiredo, and R. D. Nowak, "An EM Algorithm for Wavelet-Based Image Restoration," *IEEE Trans. Image Proc.*, vol. 8, pp. 906-916, Aug. 2003.

[17] U. Fawer and B. Aazhang, "A multiuser receiver for code division multiple access communications over multipath channels," *IEEE Trans. Commun.* vol. 43, pp. 1556-1565, Feb./Mar./Apr. 1995.

[18] S. E. Bensley and B. Aazhang, "Subspace-based blind channel estimation for code division multiple access communication systems," *IEEE Trans. Commun.* vol. 44, pp. 1009-1020, Aug. 1996.

[19] Y. Yao and V. Poor, "Eavesdropping in the synchronous CDMA channel: an EM-based approach," *IEEE Trans. Signal Proc.*, vol. 49, pp. 1748-1756, Aug. 2001.

[20] Y. Yao and V. Poor, "Blind detection of synchronous CDMA in non-Gaussian channels," *IEEE Trans. Signal Proc.*, vol. 52, pp. 271-279, Jan. 2004.

[21] N. D. Sidiropoulos, G. B. Giannakis, R. Bro, "Blind PARAFAC receivers for DS-CDMA channels," *IEEE Trans. Signal Proc.*, vol. 48, pp. 810-823, Mar. 2000.

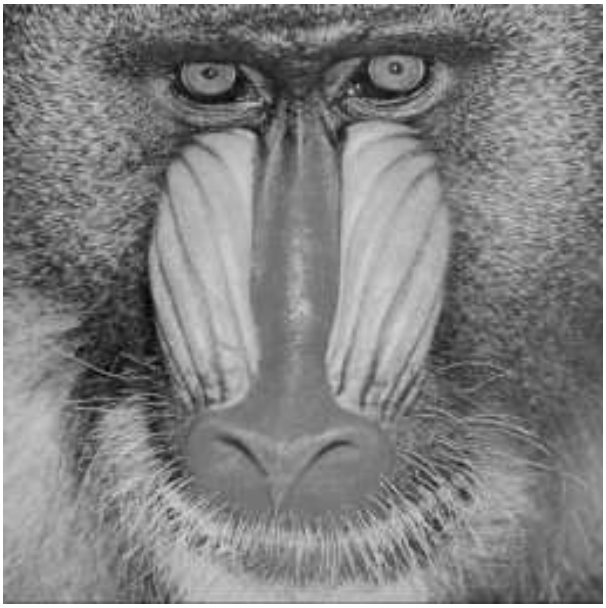
[22] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, 1988.

[23] C. D. Richmond, "PDF'S, confidence regions, and relevant statistics for a class of sample covariance-based array processors," *IEEE Trans. Signal Proc.*, vol. 44, pp. :1779-1793, Jul. 1996.

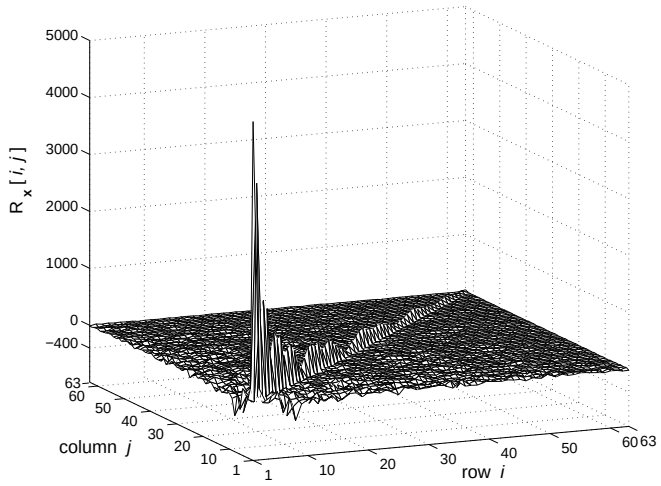
[24] M. Gkizeli, D. A. Pados, S. N. Batalama, and M. J. Medley, "Blind iterative recovery of spread-spectrum steganographic messages," in *Proc. IEEE Intern. Conf. Image Proc.*, Genoa, Italy, Oct. 2005.

[25] N. Antoniadis and A. O. Hero, "Time-delay estimation for filtered Poisson processes using an EM-type algorithm," *IEEE Trans. Signal Proc.*, vol. 42, pp. 2112-2123, Aug. 1994.

Fig. 1. F-16 image example $\mathbf{H} \in \{0, 1, \cdots, 255\}^{512\times512}$.



(a)



(b)

Fig. 2. Sample image manifesting non-white autocorrelation matrix: (a) Baboon image example $\mathbf{H} \in \{0, 1, \cdots, 255\}^{256\times256}$. (b) Host data autocorrelation matrix (8×8 DCT, 63-bin host).

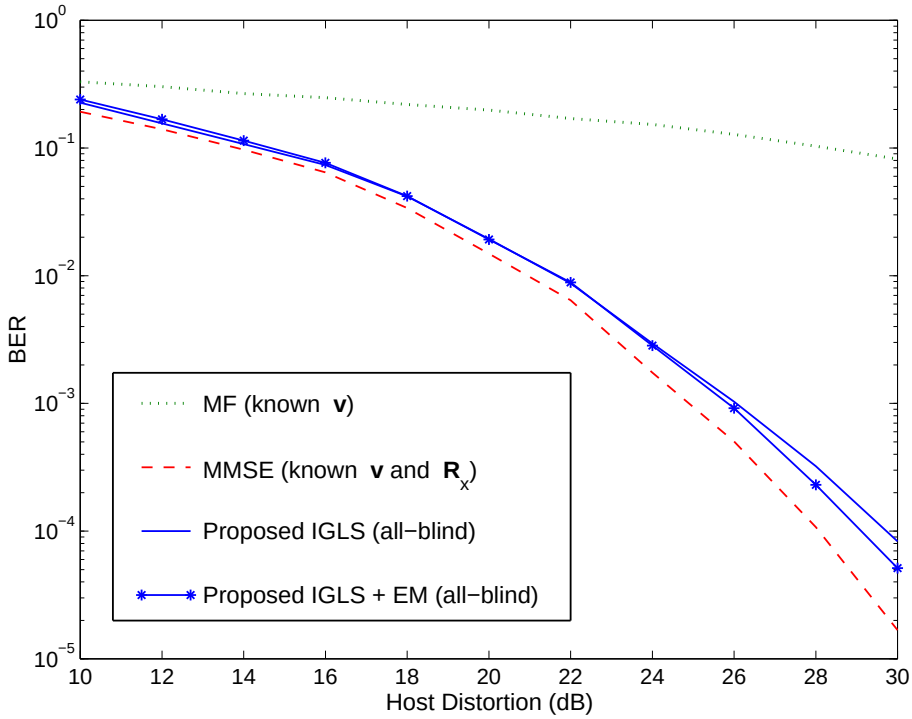


Fig. 3. Bit-error-rate versus host distortion for the 512×512 Aircraft image.

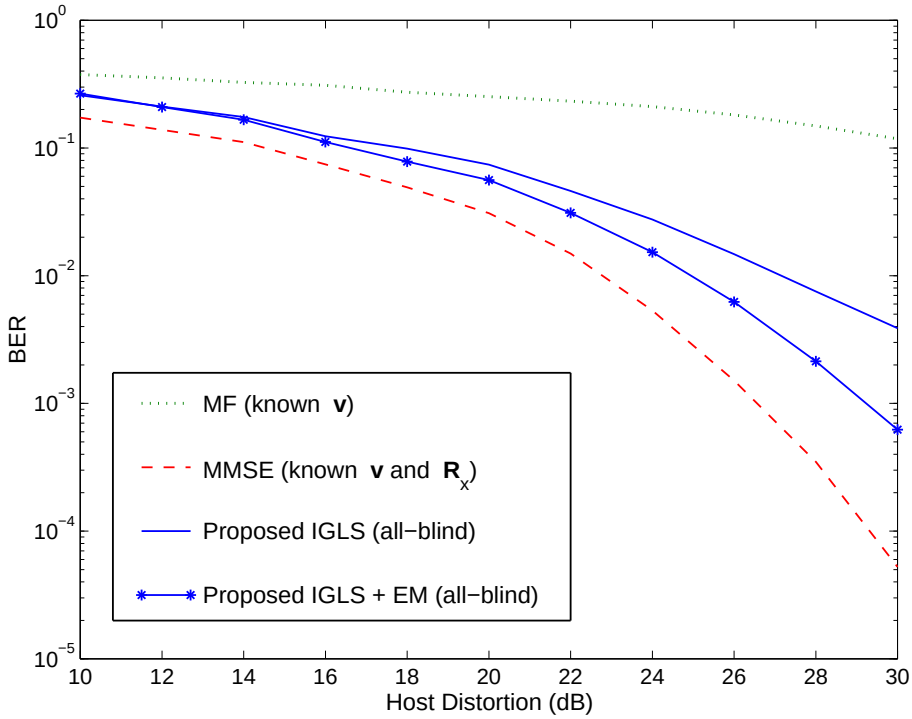


Fig. 4. Bit-error-rate versus host distortion for the 280×280 Aircraft image.

(a)

(b)

Fig. 5. A sense of distortion: (a) Original 280×280 plane image. (b) Watermarked plane (1.2 kbit hidden message, $L = 63$, distortion $\mathcal{D} = 30$ dB).

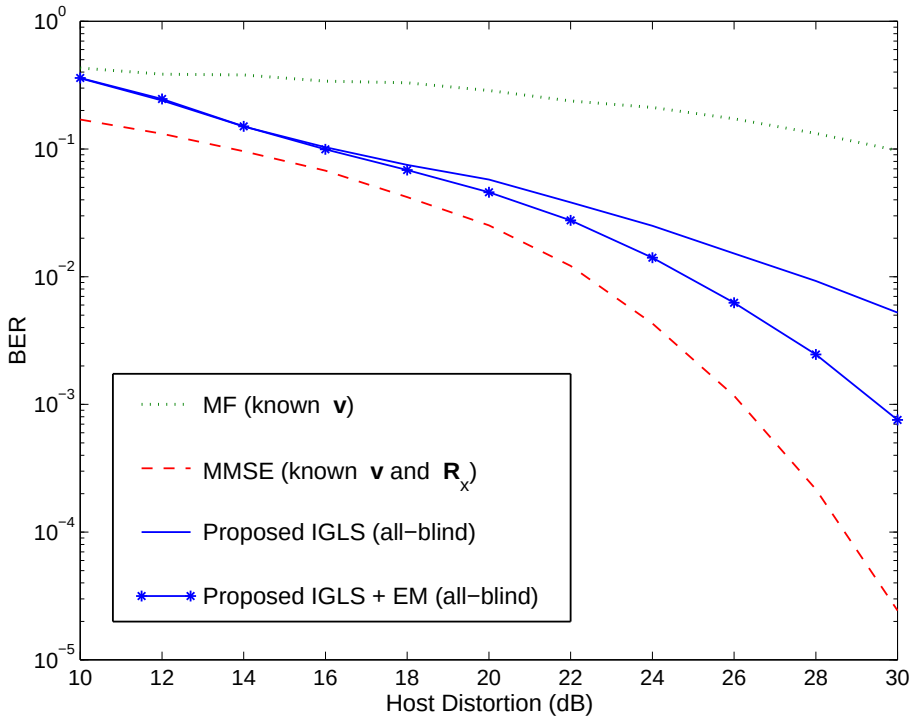


Fig. 6. Bit-error-rate versus host distortion for the 280×280 Lena image.

(a)

This is a steganalysis test. We test two algorithms for blind recovery of a message and its spreading code: iterative generalized least squares and expectation maximization.

(b)

This is a steganalysis test. We test two algorithms for blind recovery /f a m**b3**sage and its spre**e**ding cod%: it**d**rative**!**generalized least squares and expectation**0**maximization.

(c)

This is a steganalysis test. We test two algorithms for blind recovery of a message and its spreading cod**%2** iterative generalized least squares and expectation maximization.

Fig. 7. Covert messaging: (a) Original hidden ASCII message. (b) Recovered ASCII message via IGLS coupled signature and binary message estimation. (c) Recovered ASCII message via EM signature estimation (280×280 plane image, $L = 63$, distortion $\mathcal{D} = 26$ dB).